**Neuron, Volume *87***

## Supplemental Information

# Structured Variability in Purkinje Cell Activity during Locomotion

**Britton A. Sauerbrei, Evgueniy V. Lubenov, and Athanassios G. Siapas**

## Supplemental Experimental Procedures

### Task and behavioral measurements

Three male Long-Evans rats (age 3 months) were trained to walk for water reward on a 1.8m-long linear track (**Figure 1A**). Three LEDs were fixed to the headstage, and animal position was monitored using an overhead camera with video acquisition timestamped on the same global clock as electrophysiological recordings. Head pitch and roll were estimated from data acquired with a headstage-mounted inertial measurement unit using a Kalman smoother. Bipolar patch EMG electrodes (Loeb and Gans, 1986) were implanted in the acromiotrapezius muscle (English, 1978; Greene, 1968) and were differentially amplified to monitor locomotion. Electrode location was verified by postmortem dissection. Steps were identified by filtering the EMG with a 100-1000Hz finite impulse response filter, rectifying and smoothing the signals using a 100ms Gaussian kernel, then searching for sequences of peaks (local maxima) occurring during track traversal (**Figure S2**). Stepping sequences were required to include at least four peaks, with consecutive peaks separated by 220-450 ms. The step phase was obtained by linear interpolation, with the EMG peak defined as 0 for cells ipsilateral to the muscle and on the midline, and as $\pi$ for contralateral cells. All animal procedures were in accordance with the National Institutes of Health (NIH) guidelines, and with the approval of the Caltech Institutional Animal Care and Use Committee. The animals were kept on a 12-hour light cycle, followed by 12 hours of darkness, and were run in two daily sessions: one approximately four hours into the light cycle, and one approximately four hours into the dark cycle.

### Electrophysiology

Rats were chronically implanted with a 24-tetrode microdrive array. Twenty of the tetrodes targeted lobules V and VI of the cerebellar vermis, spanning +/- 1mm of the midline (**Figure S3**). Four additional tetrodes targeted the visual cortex and the hippocampus. Tetrodes were lowered to their targets over several days following implantation and subsequently adjusted in small increments to maintain unit isolation. Signals were buffered on the headstage, amplified, and acquired as 24-bit samples at 25kHz (National Instruments PXI-4498 cards) using custom LabView-based acquisition software. A screw in the skull overlying the left cerebellum was used as a reference for all signals. Spikes were clustered by fitting a mixture model in a 12-dimensional feature space (3 waveform principal components per tetrode channel) (Calabrese and Paninski, 2011; Ecker et al., 2014; Tolias et al., 2007). For cells with large and stable complex spikes, an additional stage of processing was used to distinguish complex from simple spikes. The initial, fast segment of the complex spikes had the same amplitude ratios across tetrode channels as the simple spikes for the same unit (de Solages et al., 2008), so that a single unit cluster for a Purkinje cell contained both simple and complex spikes. A matched filter for complex spikes was constructed by convolving a 25-30Hz finite impulse response filter with a decaying exponential function having a time constant of 5ms. This filter was applied to the raw, broadband data, and the peaks in the resulting signal were registered with the preceding spike from the corresponding unit cluster. The broadband waveforms were extracted within -2ms and 10ms of these putative complex spike events and sorted using principal component features. Thus, each spike in the unit cluster was identified as either simple or complex. For analyses that did not compare simple and complex spike activity, the cell's spike train was taken to be the union of spikes of both types. At the end of each

experiment, electrolytic lesions were applied at each recording site, and the tetrode locations were verified in Hoescht-stained tissue sections (**Figure S3**).

**Statistical analysis**

The state-dependent mean firing rates for locomotion, licking, and inactivity were defined as the total number of spikes occurring in the state divided by the total duration of the state throughout the dataset. For each Purkinje cell, a circular distribution was fit to the step phases sampled at the spike times. A second-order generalized von Mises distribution (Gatto and Jammalamadaka, 2007) was fit for cells with one or two peaks in the step cycle. For cells with three peaks, a kernel density estimator was used with a wrapped normal distribution of width $(4/3N)^{0.2}v$, where v is the sample circular standard deviation (Mardia and Jupp, 2000). To determine whether spiking was locked to the step cycle, Kuiper's test was used (Mardia and Jupp, 2000) for both simple and complex spikes. Lick modulation was determined using the same procedure, with the lick onset defined to be 0 degrees.

The step-to-step variability of spike counts was quantified using the variance-to-mean ratio, or Fano factor (**Figure 2D**). For each neuron and window length Δt, the Fano factor was computed as:

$$FF(\Delta t) = \frac{var(n(\Delta t))}{mean(n(\Delta t))}$$

where n(Δt) is the number of spikes lying within a window of length Δt ms following the EMG peak for each step.

The shape of the step-locked firing rate curves was investigated by smoothing the spike trains with a 25ms Gaussian kernel and extracting the curves for each step, with the time axis normalized so that each curve was parameterized by step phase θ, rather than time. Principal component analysis was performed on the curves for each cell, and the effect of each component, w(θ), was visualized as a perturbation of the mean curve, f(θ) (Ramsay and Silverman, 2005). The extent to which each component was an additive shift, a multiplicative scaling, or a phase shift was determined by computing bias, amplitude, and phase scores, respectively. A pure bias shift corresponds to a completely flat component, so the bias score was defined to be:

$$S_{bias} = |w(\theta) \bullet v(\theta)|$$

with $v(\theta) = 1/\sqrt{n}$, where n = 50 is the length of the curve. A pure shift in amplitude corresponds to a component with the same shape as the mean curve, but rescaled by a multiplicative constant. The amplitude score was defined using:

$$f^*(\theta) = \frac{f(\theta) - \mu_f}{\|f(\theta) - \mu_f\|}$$

$$S_{amp} = |w(\theta) \bullet f^*(\theta)|$$

where $\mu_f$ is the mean of f(θ). A pure phase shift of δ ≠ 0 would correspond to a transformation of f(θ) to f(θ + δ), so the phase score was defined using:

$$g_\delta(\theta) = \frac{f(\theta+\delta) - f(\theta)}{\|f(\theta+\delta) - f(\theta)\|}$$

$$S_{phase} = \max_{\delta \in (-\Delta, \Delta) \setminus \{0\}} |w(\theta) \bullet g_\delta(\theta)|$$

where Δ = 22 degrees. The range of possible offsets was limited to 2Δ in order to regularize the phase score estimation.

The influence of behavioral variables on Purkinje cell activity across steps was first determined by averaging speed, forward acceleration, sin(roll), and sin(pitch), and EMG amplitude within each step cycle. Larger values of pitch reflect upward movement of the head (dorsiflexion). Positive values of roll correspond to rightward rotations of the head for cells on the right side, and to leftward rotation for cells on the left side and midline. To study the variation in curve shape with behavior, slicing intervals for behavioral values were determined using an equal-count algorithm with twelve (**Figure 4B**) or six (**Figure 4C**) intervals and 50% overlap (Cleveland, 1993). Step-locked firing rates were then averaged within each interval.

In order to quantify the influence of behavior on spiking, we defined the firing rate on each step to be the number of spikes occurring during that step divided by the step duration. We converted the firing rate and the behavioral variables to z-scores, and estimated a multiple regression model for each cell:

$$Z_{rate} = \beta_{speed} Z_{speed} + \beta_{acc} Z_{acc} + \beta_{roll} Z_{roll} + \beta_{pitch} Z_{pitch} + \beta_{EMG} Z_{EMG} + \epsilon$$

where $Z_{rate}$ is the z-score of the step-locked firing rates, $Z_{roll}$ is the z-score of sin(roll), $Z_{pitch}$ is the z-score of sin(pitch), $Z_{EMG}$ is the z-score of the EMG amplitude, and ε is a normally-distributed error term. The normality of the residuals was checked using quantile-quantile plots (**Figure S7A**). Performing the same analysis using rank-transformed data, substituting the z-scored ranks of the independent and dependent variables for the original linear data, produced nearly identical parameter estimates (**Figure S7B**).

In order to study how the relationship between neural activity and behavior varies throughout the step cycle, we next estimated step-phase-dependent regression curves (**Figure 4D**). For each step phase θ, we estimated the following model:

$$Z_{rate}(\theta) = \beta_{speed}(\theta) Z_{speed}(\theta) + \beta_{acc}(\theta) Z_{acc}(\theta) + \beta_{roll}(\theta) Z_{roll}(\theta) + \beta_{pitch}(\theta) Z_{pitch}(\theta) + \beta_{EMG}(\theta) Z_{EMG}(\theta) + \epsilon$$

where $Z_{rate}(\theta)$ is firing rate at phase θ, z-scored with respect to the distribution of firing rates at θ, and $Z_{speed}(\theta)$ is the z-scored speed at θ.

We characterized the shape of these regression curves by computing scores for bias ($S_{bias}$), amplitude ($S_{amp}$), and phase ($S_{phase}$), by substituting $\beta_{var}(\theta) / \|\beta_{var}(\theta)\|$ for w(θ) for each

behavioral variable. This produced an ordered triplet ($S_{bias}$, $S_{amp}$, $S_{phase}$) for each significantly tuned cell ($\beta_{var} \neq 0$ at q = .05). For each behavioral variable, we performed an agglomerative hierarchical clustering analysis of these triplets using a Ward linkage function and a gap criterion for determining the number of clusters (Everitt, 2011) (**Figure 4E**).

Coordinated activity between pairs of cells was assessed using the partial rank correlations in the firing rates: the effects of the behavioral variables were first removed using the multiple regression model, and Spearman's ρ was computed between the residuals. Partial rank correlations computed using the residuals from the rank-transformed regression model produced nearly identical correlations (**Figure S7C**).

The low rate of complex spikes did not permit the analysis of step-locked firing rates; instead, we studied the effects of behavioral variables on the number of complex spikes, C, within a 350ms window starting at the onset of each step cycle. For each Purkinje cell with stable complex spikes, we estimated a Poisson regression model:

$$log(E(C|Z_{speed}, Z_{roll}, Z_{pitch})) = \mu_C + \beta_{speed}Z_{speed} + \beta_{acc}Z_{acc} + \beta_{roll}Z_{roll} + \beta_{pitch}Z_{pitch} + \beta_{EMG}Z_{EMG} + \epsilon$$

where $\mu_C$ is the mean number of spikes, and $Z_{speed}$, $Z_{acc}$, $Z_{roll}$, $Z_{pitch}$, $Z_{EMG}$ and ε are defined as in the previous equation.

For the analysis of step phase and lick modulation, the linear models, and the pairwise correlations, corrections for multiple comparisons were made by setting the false discovery rate to q = .05 (Benjamini and Hochberg, 1995). Data from all animals, Purkinje cells, and steps were included in the analysis. For parametric tests, the assumption of normality was checked using normal quantile-quantile plots.

## Supplemental References

de Solages, C., Szapiro, G., Brunel, N., Hakim, V., Isope, P., Buisseret, P., Rousseau, C., Barbour, B., and Léna, C. (2008). High-frequency organization and synchrony of activity in the purkinje cell layer of the cerebellum. Neuron *58*, 775-788.

English, A.W. (1978). An electromyographic analysis of forelimb muscles during overground stepping in the cat. The Journal of experimental biology *76*, 105-122.

Everitt, B. (2011). Cluster analysis, 5th edn (Chichester, West Sussex, U.K.: Wiley).

Gatto, R., and Jammalamadaka, S.R. (2007). The generalized von Mises distribution. Statistical Methodology *4*, 341-353.

Greene, E. (1968). Anatomy of the rat (New York: Hafner Pub. Co.).

Loeb, G.E., and Gans, C. (1986). Electromyography for experimentalists (Chicago: University of Chicago Press).

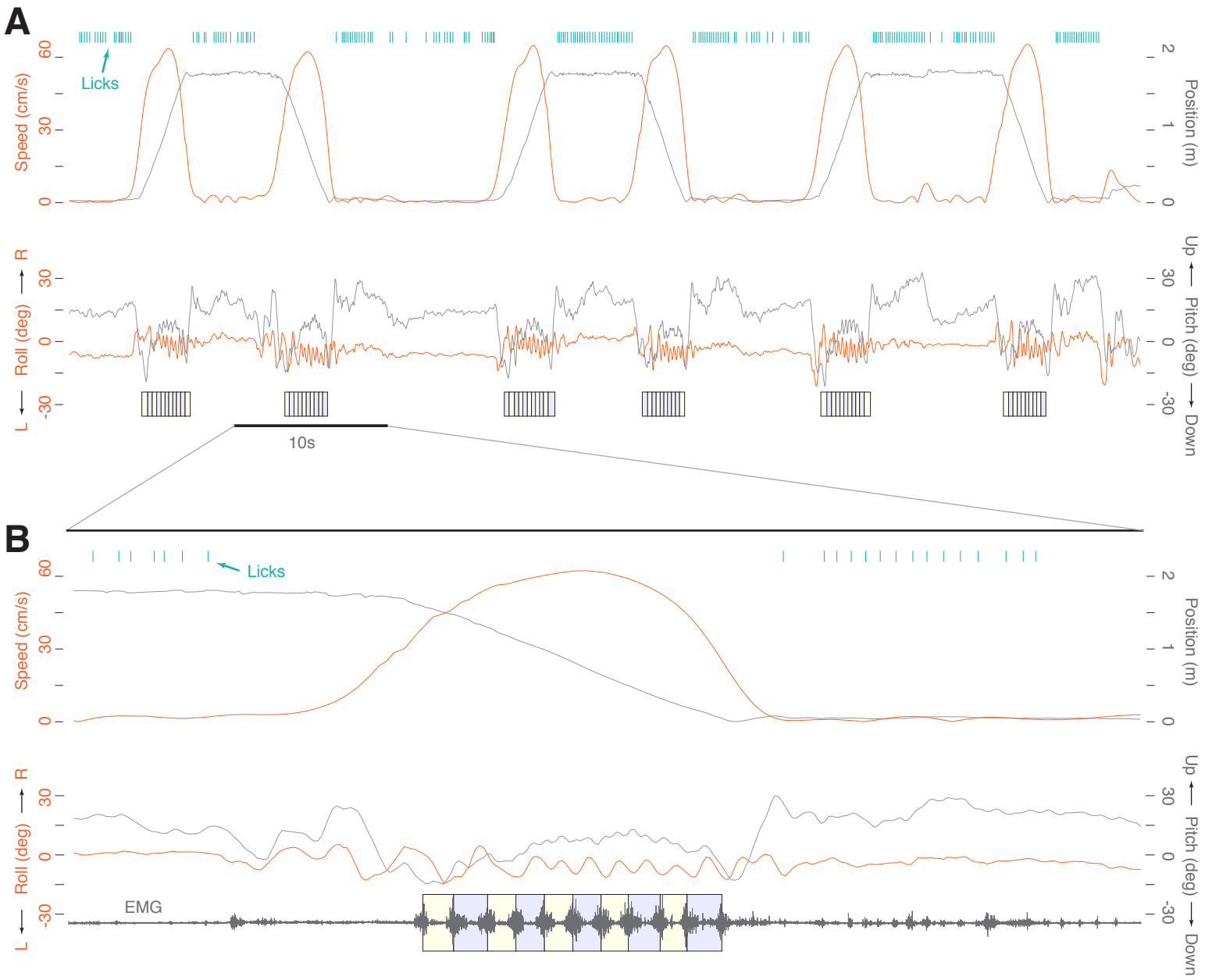


**Figure S1, related to Figure 1: Track behavior.** (**A**) Animal speed (orange) and position (gray) over the course of six laps, along with head roll (orange), head pitch (gray), lick times (teal), and step segmentation (yellow and blue rectangles). (**B**) A ten-second segment of the same data showing a single track traversal, including the raw EMG trace.

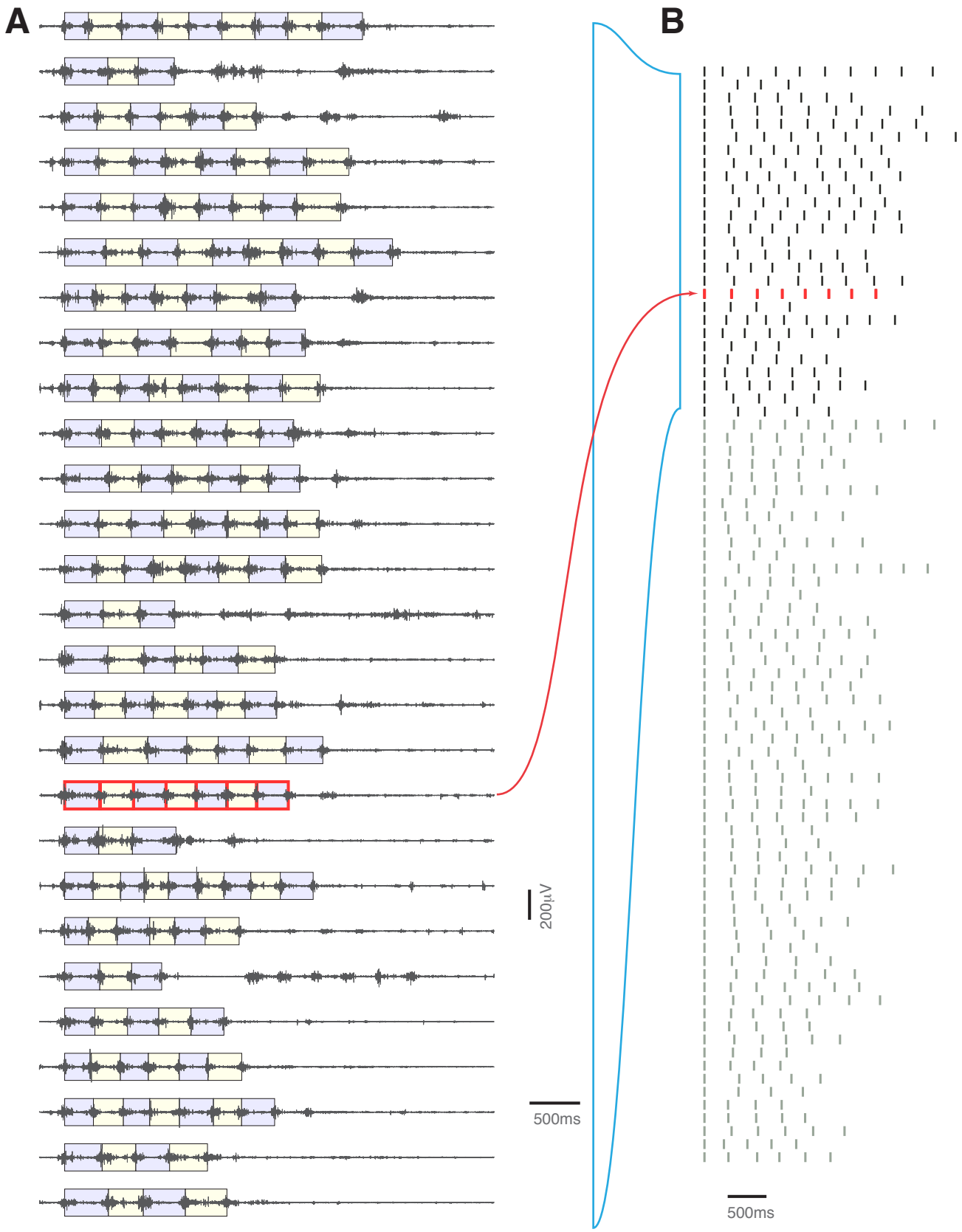


**Figure S2, related to Experimental Procedures:** Step segmentation from EMG. (**A**) Example EMGs and step partitioning. (**B**) Raster plot of EMG peaks for the full dataset. The example segment from **fig. 1C** is highlighted in red.

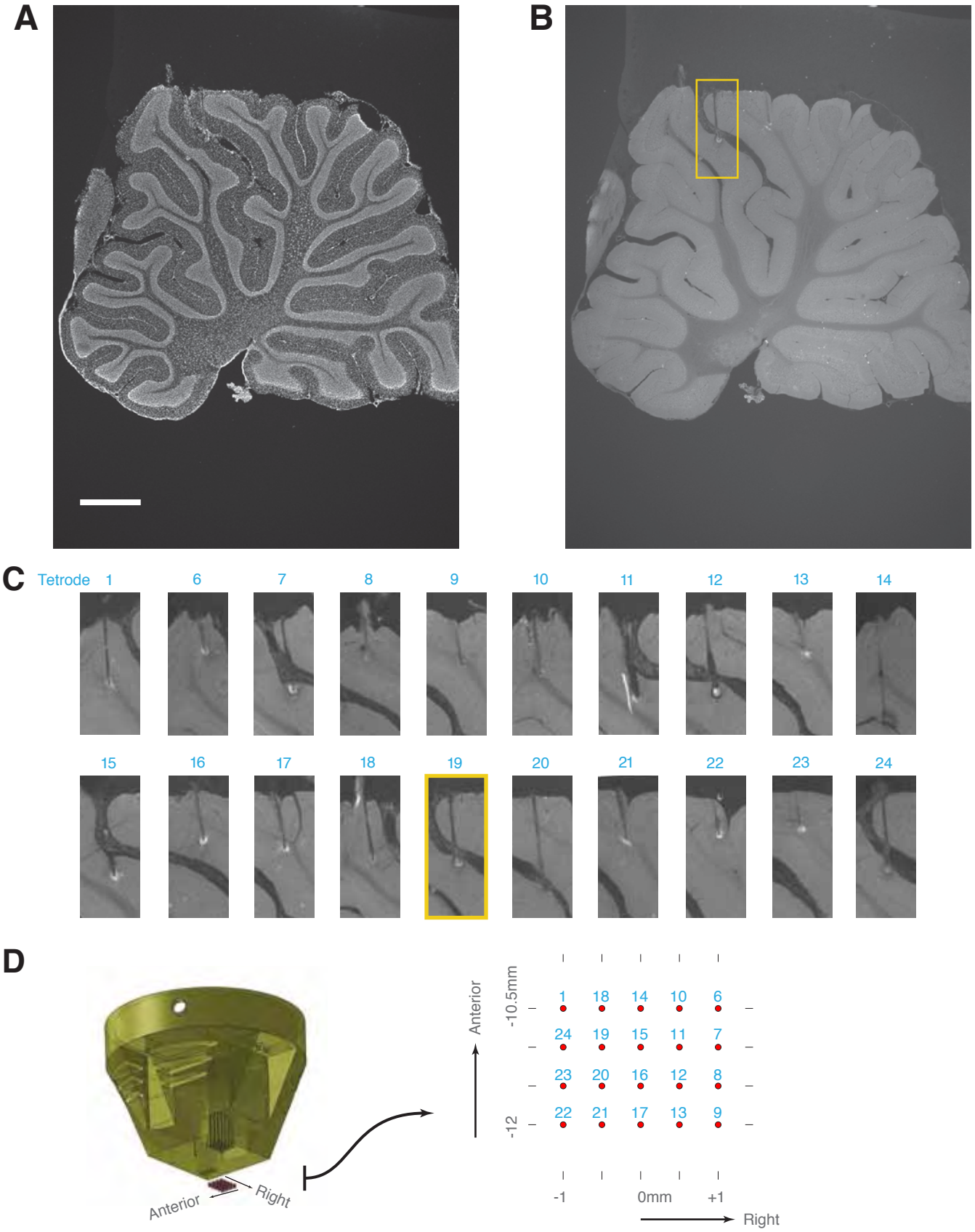


**Figure S3, related to Experimental Procedures:** Histological verification of tetrode recording locations. (**A**) Hoechst-stained sagittal section showing cerebellar lobules and cortical layers. Scale bar: 1mm. Peak of excitation spectrum: 350nm. (**B**) The same section imaged using 480nm excitation. The tetrode track and electrolytic lesion for tetrode 19 are clearly visible. (**C**) Recording locations for all tetrodes. 480nm excitation. (**D**) Design and stereotactic coordinates for custom-built multi-tetrode array.

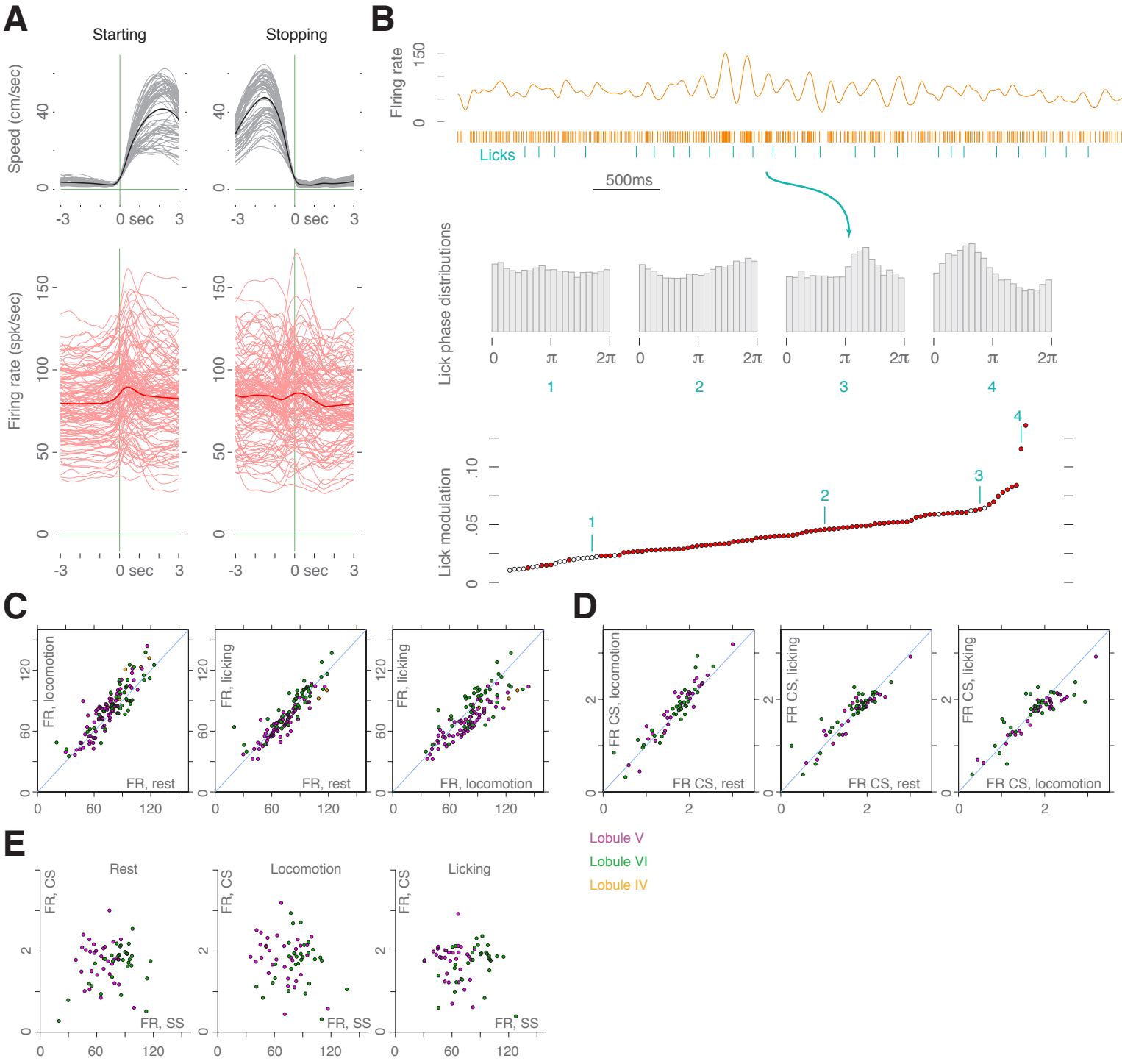


**Figure S4, related to Figure 1:** Purkinje cell activity during the initiation and termination of movement and during licking. (**A**) Above: animal speed centered on the beginning (left) and end (right) of laps. Each gray curve is the average speed for a single dataset, and the black curves are the averages. Below: Purkinje cell firing rates centered on the beginning and end of laps. Each light red curve is a single cell, and the bold red lines are averages. Most cells exhibit a large phasic increase in firing rate at movement onset and a smaller tonic increase during locomotion. Many also show phasic increases or decreases in firing rate during stopping. Spike trains were smoothed with a 200ms Gaussian kernel. (**B**) Many Purkinje cells discharge rhythmically during licking. Top: example spike raster for a single licking bout. Middle: lick phase histograms for four example Purkinje cells. The three cells on the right are significantly tuned to lick phase. Bottom: distribution of lick tuning values (Kuiper's statistic) for all Purkinje cells. Significantly tuned cells are red. (**C**) State-dependent firing rates for pairs of states. For most cells, firing rates are elevated during locomotion, relative to rest and licking. (**D**) State-dependent complex spike rates for pairs of states. Rates were higher during locomotion than rest and licking. (**E**) State-dependent complex versus simple spike firing rates. Simple and complex spike rates are not correlated for any state.

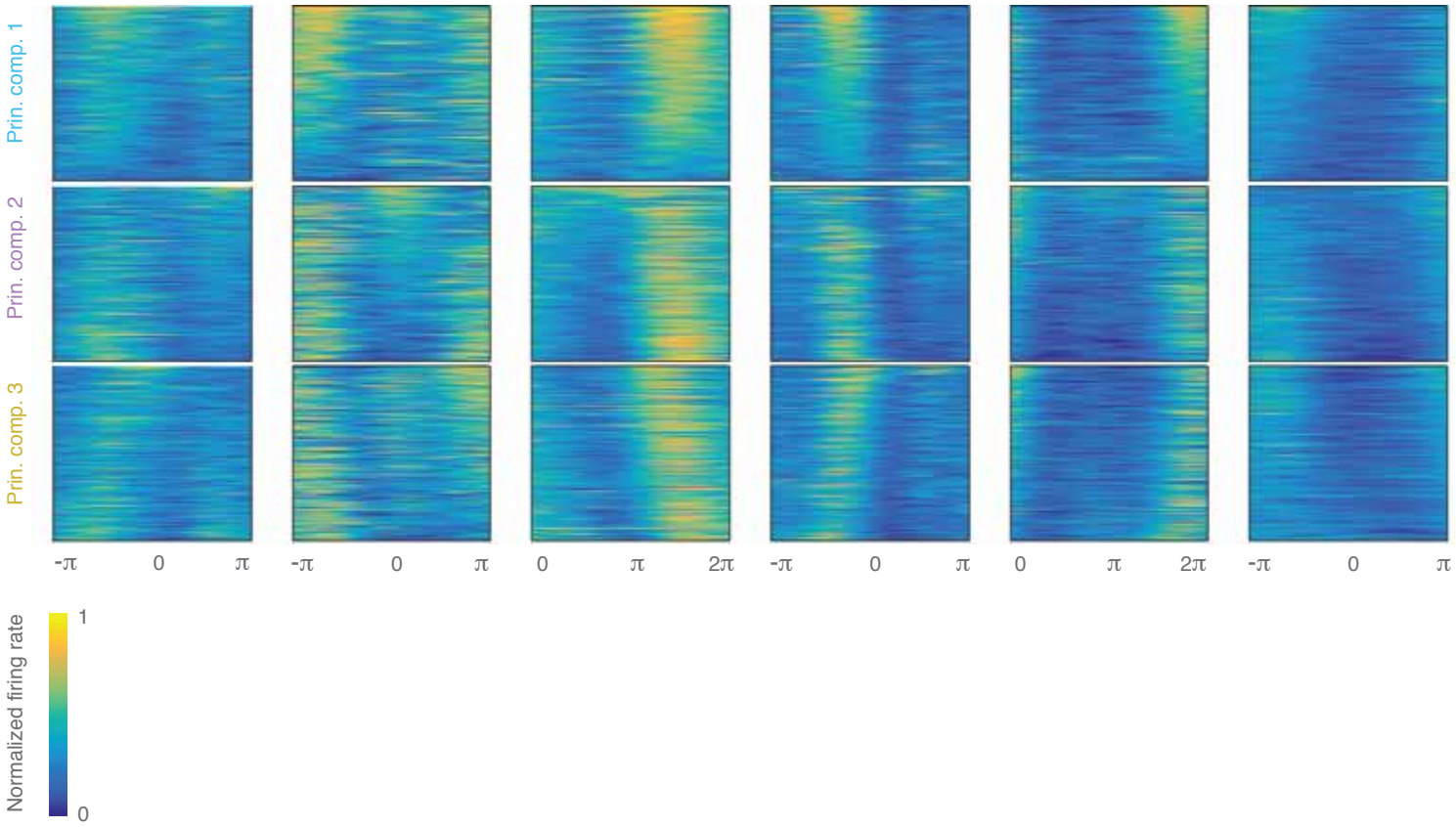


**Figure S5, related to Figure 3:** firing rates on all step cycles for the cells in fig. 3, ordered by scores for the first three principal components.

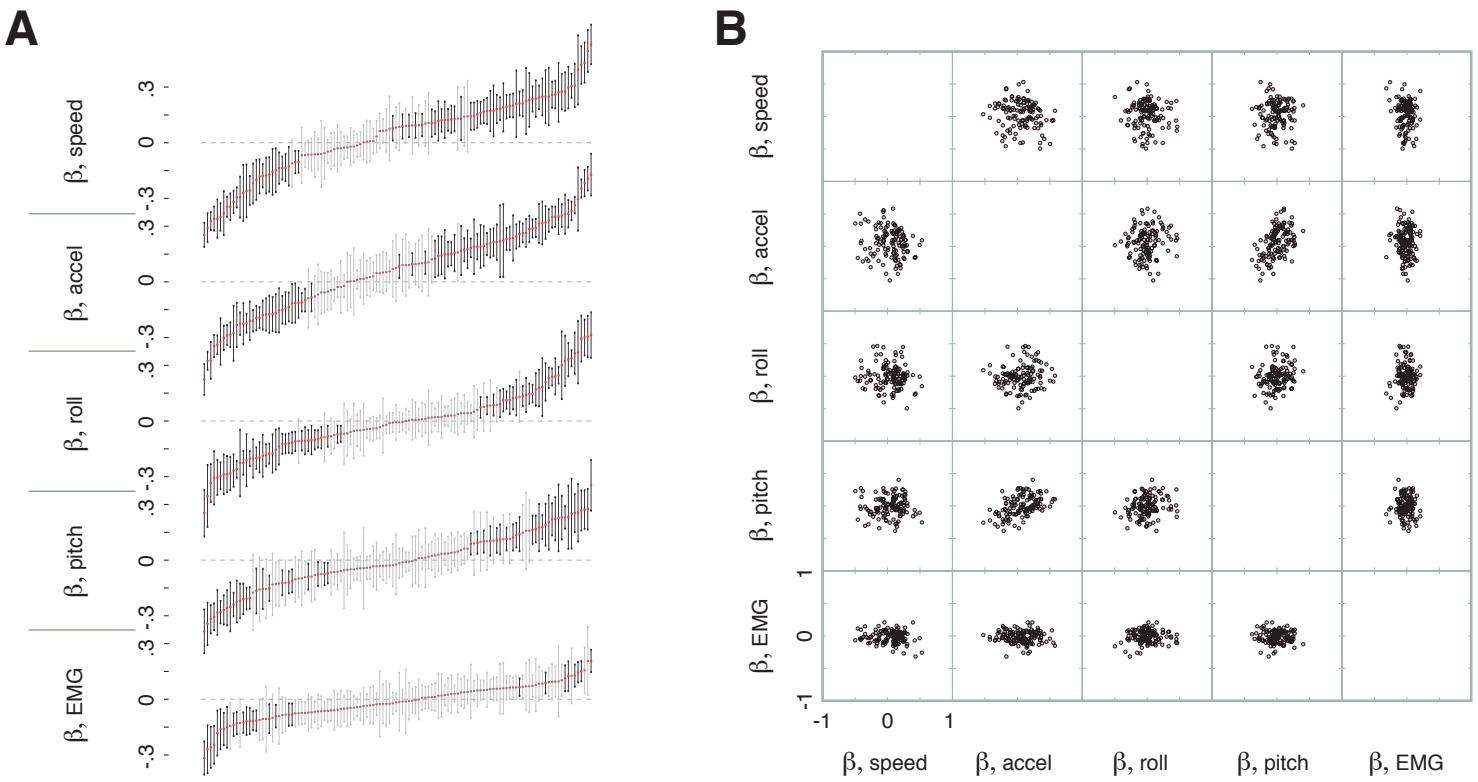


**Figure S6, related to Figure 4:** regression of step-locked average firing rates against step-averaged behavioral variables. (**A**) Regression coefficients and 95% confidence intervals for all cells, sorted by coefficient value. Coefficients significantly different from zero are shown in black, with the coefficients not reaching significance in gray. All variables are Z-scores. Most Purkinje cells are modulated by speed, acceleration, or head posture across steps. (**B**) Scatterplots of regression coefficient pairs. Each subplot corresponds to a pair of behavioral variables, and each point represents the pair of regression coefficients for a single cell.

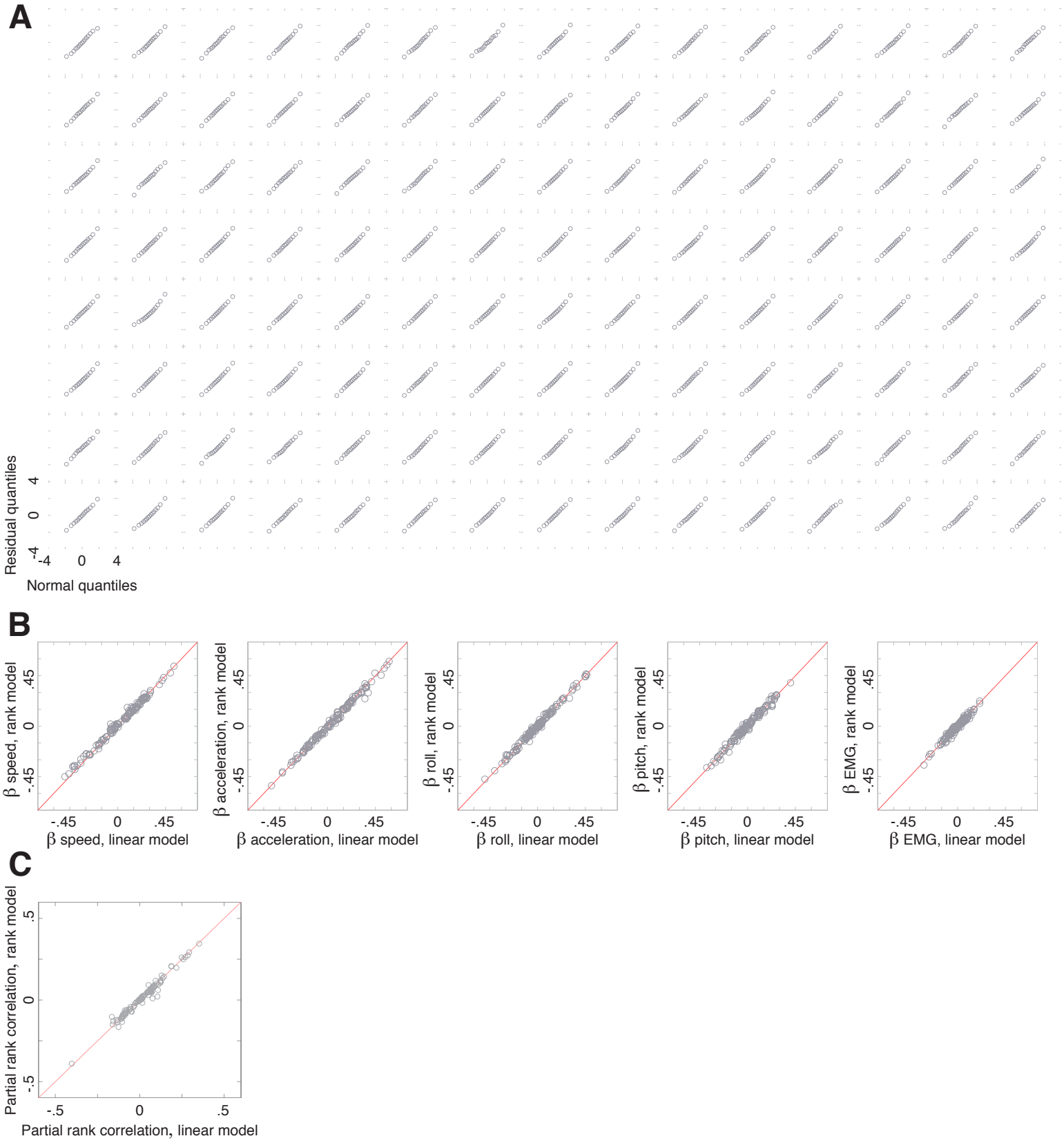


**Figure S7, related to Figure 5:** Linear regression yields normally-distributed residuals and a model invariant to rank transformation. (**A**) Normal quantile-quantile plots for the regression residuals for all 120 Purkinje cells. Each panel plots the quantiles of the regression residuals for one cell against the quantiles of a normal distribution with the same mean and standard deviation. The points fall along a straight diagonal line, indicating that the residuals are normally distributed. (**B**) Model coefficients for rank regression versus linear regression. (**C**) Partial rank correlations between pairs of cells using the residuals from the rank regression versus the linear regression.